CRYSTALLIZATION OF DIPEPTIDYL PEPTIDASE IV (DPPIV)

RELATED APPLICATION

[001] This application claims the benefit of U.S. Provisional Application No. 60/409,206, filed September 9, 2002, which is incorporated herein by reference.

FIELD OF THE INVENTION

[002] The present invention relates to a member of the S9 family of human proteases known as Dipeptidyl Peptidases (DPP) and more specifically to a particular dipeptidyl peptidase known as dipeptidyl peptidase IV (DPPIV). Provided is DPPIV in crystalline form, methods of forming crystals comprising DPPIV, methods of using crystals comprising DPPIV, a crystal structure of DPPIV, and methods of using the crystal structure.

BACKGROUND OF THE INVENTION

[003] A general approach to designing inhibitors that are selective for a given protein is to determine how a putative inhibitor interacts with a three dimensional structure of that protein. For this reason it is useful to obtain the protein in crystalline form and perform X-ray diffraction techniques to determine the protein's three-dimensional structure coordinates. Various methods for preparing crystalline proteins are known in the art.

[004] Once protein crystals are produced, crystallographic data can be generated using the crystals to provide useful structural information that assists in the design of small molecules that bind to the active site of the protein and inhibit the protein's activity *in vivo*. If the protein is crystallized as a complex with a ligand, one can determine both the shape of the protein's binding pocket when bound to the ligand, as well as the amino acid residues that are capable of close contact with the ligand. By knowing the shape and amino acid residues comprised in the binding pocket, one may design new ligands that will interact favorably with the protein. With such structural information, available computational methods may be used to predict how strong the ligand binding interaction will be. Such methods aid in the design of inhibitors that bind strongly, as well as selectively to the protein.

Express Mailing No. EL978337872US

SUMMARY OF THE INVENTION

[005] The present invention is directed to crystals comprising DPPIV and particularly crystals comprising DPPIV that have sufficient size and quality to obtain useful information about the structural properties of DPPIV and molecules or complexes that may associate with DPPIV.

[006] In one embodiment, a composition is provided that comprises a protein in crystalline form wherein the protein has 65%, 70, 80, 90, 95% or greater identity with residues 51-778 of SEQ. ID No. 1.

[007] In one variation, the protein has activity characteristic of DPPIV. For example, the protein may optionally be inhibited by inhibitors of wild type DPPIV.

[008] The protein may also diffract X-rays for a determination of structure coordinates to a resolution of 4 Å, 3 Å, 2.5 Å, 2 Å or less.

[009] In one variation, the protein crystal has a crystal lattice in a $P2_1$ space group. The protein crystal may also have a crystal lattice having unit cell dimensions, +/- 5%, of a=121.53Å b=124.11Å and c=144.42Å, $\alpha=\gamma=90^{o}$, $\beta=114.6^{o}$.

[0010] The present invention is also directed to crystallizing DPPIV. The present invention is also directed to the conditions useful for crystallizing DPPIV. It should be recognized that a wide variety of crystallization methods can be used in combination with the crystallization conditions to form crystals comprising DPPIV including, but not limited to, vapor diffusion, batch, and dialysis.

[0011] In one embodiment, a method is provided for forming crystals of a protein comprising: forming a crystallization volume comprising: a protein that has at least 65%, 70, 80, 90, 95% identity with residues 51-778 of SEQ. ID No. 1 in a concentration between 1 mg/ml and 50 mg/ml; 5-50% w/v of precipitant wherein the precipitant comprises one or more members of the group consisting of PEG MME having a molecular weight range between 300-10000, and PEG having a molecular weight range between 100-10000; optionally 0.05 to 0.8M additives wherein the additives comprises sarcosine or 0.5 to 25% additives wherein the additives comprises xylitrol; and wherein the crystallization volume has a pH between pH 5 and pH 9; and storing the crystallization volume under conditions suitable for crystal formation. The method optionally further comprises using 0.05-0.2M buffers selected from the group

consisting of Tris-HCl, bicine and combinations thereof. The method also optionally further includes performing the crystallization at a temperature between 1°C - 25°C.

[0012] The method may optionally further comprise forming a protein crystal that has a crystal lattice in a $P2_1$ space group. The method also optionally further comprises forming a protein crystal that has a crystal lattice having unit cell dimensions, +/- 5%, of a=121.53Å b=124.11Å and c=144.42Å, $\alpha=\gamma=90^\circ$, $\beta=114.6^\circ$. The invention also relates to protein crystals formed by these methods.

[0013] The present invention is also directed to structure coordinates for DPPIV as well as structure coordinates that are comparatively similar to these structure coordinates. It is noted that these comparatively similar structure coordinates may encompass proteins with similar sequences and/or structures, such as other members of the S9 protease family. For example, machine-readable data storage media is provided having data storage material encoded with machine-readable data that comprises structure coordinates that are comparatively similar to the structure coordinates of DPPIV. The present invention is also directed to a machine readable data storage medium having data storage material encoded with machine readable data, which, when read by an appropriate machine, can display a three dimensional representation of all or a portion of a structure of DPPIV or a model that is comparatively similar to the structure of all or a portion of DPPIV.

[0014] In one embodiment, machine readable data storage medium is provided having data storage material encoded with machine readable data, the machine readable data comprising: structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on alpha-carbon atoms positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues 51-778 of SEQ. ID No. 1.

[0015] In another embodiment, machine readable data storage medium is provided having data storage material encoded with machine readable data, the machine readable data comprising: structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those

alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in Tables 1, 2, 3 and/or 4.

[0016] It is noted in regard to these embodiments that the root mean square deviation calculation may optionally be based on a comparison of main-chain atoms, non-hydrogen atoms or a comparison of all atoms where the same type of amino acid residue is present. Also, the root mean square deviation of alpha-carbon atoms, main-chain atoms, non-hydrogen atoms or all atoms may optionally be less than 2.7 Å, 2.5 Å, 2.0 Å, 1.5 Å, 1 Å, 0.5 Å, or less.

[0017] The present invention is also directed to a three-dimensional structure of all or a portion of DPPIV. This three-dimensional structure may be used to identify binding sites, to provide mutants having desirable binding properties, and ultimately, to design, characterize, or identify ligands capable of interacting with DPPIV. Ligands that interact with DPPIV may be any type of atom, compound, protein or chemical group that binds to or otherwise associates with the protein. Examples of types of ligands include natural substrates for DPPIV, inhibitors of DPPIV, and heavy atoms.

[0018] In one embodiment, a method is provided for displaying a three dimensional representation of a structure of a protein comprising: taking machine readable data comprising structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues 51-778 of SEQ. ID No. 1; computing a three dimensional representation of a structure based on the structure coordinates; and displaying the three dimensional representation.

[0019] In another embodiment, a method is provided for displaying a three dimensional representation of a structure of a protein comprising: displaying a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates of Figure 3 that are present in residues shown in Tables 1, 2, 3 and/or 4 or residues 51-778 of SEQ. ID No. 1.

[0020] It is again noted in regard to these embodiments that the root mean square deviation calculation may optionally be based on a comparison of main-chain atoms, non-hydrogen atoms or a comparison of all atoms where the same type of amino acid residue is present. Also, the root mean square deviation of alpha-carbon atoms, non-hydrogen atoms or all atoms may optionally be less than 2.7 Å, 2.5 Å, 2.0 Å, 1.5 Å, 1 Å, 0.5 Å, or less.

[0021] The present invention is also directed to a method for solving a three-dimensional crystal structure of a target protein using the structure of DPPIV.

[0022] In one embodiment, a computational method is provided comprising: taking machine readable data comprising structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues 51-778 of SEQ. ID No. 1; computing phases based on the structural coordinates; computing an electron density map based on the computed phases; and determining a three-dimensional crystal structure based on the computed electron density map.

[0023] In another embodiment, a computational method is provided comprising: taking an X-ray diffraction pattern of a crystal of the target protein; and computing a three-dimensional electron density map from the X-ray diffraction pattern by molecular replacement, wherein structure coordinates used as a molecular replacement model comprise structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues 51-778 of SEQ. ID No. 1. This method may optionally further comprise determining a three-dimensional crystal structure based upon the computed three-dimensional electron density map.

[0024] It is again noted in regard to these embodiments that the root mean square deviation calculation may optionally be based on a comparison of main-chain atoms, non-hydrogen atoms or a comparison of all atoms where the same type of amino acid residue is

present. Also, the root mean square deviation of alpha-carbon atoms, main-chain atoms, non-hydrogen atoms or all atoms may optionally be less than 2.7 Å, 2.5 Å, 2.0 Å, 1.5 Å, 1 Å, 0.5 Å, or less.

[0025] The present invention is also directed to using a crystal structure of DPPIV, in particular the structure coordinates of DPPIV and the surface contour defined by them, in methods for screening, designing, or optimizing molecules or other chemical entities that interact with and preferably inhibit DPPIV.

[0026] One skilled in the art will appreciate the numerous uses of the inventions described herein, particularly in the areas of drug design, screening and optimization of drug candidates, as well as in determining additional unknown crystal structures. For example, a further aspect of the present invention relates to using a three-dimensional crystal structure of all or a portion of DPPIV and/or its structure coordinates to evaluate the ability of entities to associate with DPPIV. The entities may be any entity that may function as a ligand and thus may be any type of atom, compound, protein (such as antibodies) or chemical group that can bind to or otherwise associate with a protein.

[0027] In one embodiment, a method is provided for evaluating a potential of an entity to associate with a protein comprising: creating a computer model of a protein structure using structure coordinates that comprise structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues 51-778 of SEQ. ID No. 1; performing a fitting operation between the entity and the computer model; and analyzing results of the fitting operation to quantify an association between the entity and the model.

[0028] In another embodiment, a method is provided for evaluating a potential of an entity to associate with a protein comprising: computing a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3 Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates that are present in residues shown in Tables

1, 2, 3 and/or 4 or residues 51-778 of SEQ. ID No. 1; evaluating a potential of an entity to associate with the surface contour by performing a fitting operation between the entity and the surface contour; and analyzing results of the fitting operation to quantify an association between the entity and the computer model.

[0029] In another embodiment, a method is provided for identifying entities that can associate with a protein comprising: generating a three-dimensional structure of a protein using structure coordinates that comprise structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues 51-778 of SEQ. ID No. 1; employing the three-dimensional structure to design or select an entity that can associate with the protein; and contacting the entity with a protein having at least 65% identity with residues 51-778 of SEQ. ID No. 1.

[0030] In another embodiment, a method is provided for identifying entities that can associate with a protein comprising: computing a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates that are present in residues shown in Tables 1, 2, 3 and/or 4 or residues 51-778 of SEQ. ID No. 1; employing the computer model to design or select an entity that can associate with the protein; and contacting the entity with a protein having at least 65%, 70, 80, 90, 95% identity with residues 51-778 of SEQ. ID No. 1.

[0031] In another embodiment, a method is provided for evaluating the ability of an entity to associate with a protein, the method comprising: constructing a computer model defined by structure coordinates that comprise structure coordinates that have a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on alpha-carbon atom positions of corresponding atomic coordinates of Figure 3, the root mean square deviation being calculated based only on those alpha-carbon atoms of amino acid residues in the structure coordinates that are also present in residues shown in Tables 1, 2, 3 and/or 4 or residues 51-778

of SEQ. ID No. 1; selecting an entity to be evaluated by a method selected from the group consisting of (i) assembling molecular fragments into the entity, (ii) selecting an entity from a small molecule database, (iii) *de novo* ligand design of the entity, and (iv) modifying a known ligand for DPPIV, or a portion thereof; performing a fitting program operation between computer models of the entity to be evaluated and the binding pocket in order to provide an energy-minimized configuration of the entity in the binding pocket; and evaluating the results of the fitting operation to quantify the association between the entity and the binding pocket model in order to evaluate the ability of the entity to associate with the binding pocket.

[0032] In another embodiment, a method for evaluating the ability of an entity to associate with a protein, the method comprising: computing a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates that are present in residues shown in Tables 1, 2, 3 and/or 4 or residues 51-778 of SEQ. ID No. 1; selecting an entity to be evaluated by a method selected from the group consisting of (i) assembling molecular fragments into the entity, (ii) selecting an entity from a small molecule database, (iii) *de novo* ligand design of the entity, and (iv) modifying a known ligand for an DPPIV, or a portion thereof; performing a fitting program operation between computer models of the entity to be evaluated and the binding pocket in order to provide an energy-minimized configuration of the entity in the binding pocket; and evaluating the results of the fitting operation to quantify the association between the entity and the binding pocket model in order to evaluate the ability of the entity to associate with the said binding pocket.

[0033] It is again noted in regard to these embodiments that the root mean square deviation calculation may optionally be based on a comparison of main-chain atoms, non-hydrogen atoms or a comparison of all atoms where the same type of amino acid residue is present. Also, the root mean square deviation of alpha-carbon atoms, non-hydrogen atoms or all atoms may optionally be less than 2.7 Å, 2.5 Å, 2.0 Å, 1.5 Å, 1 Å, 0.5 Å, or less.

[0034] Also in regard to each of these embodiments, the protein may optionally have activity characteristic of DPPIV. For example, the protein may optionally be inhibited by inhibitors of wild type DPPIV.

[0035] In another embodiment, a method is provided for identifying an entity that associates with a protein comprising: taking structure coordinates from diffraction data obtained from a crystal of a protein that has at least 65%, 70%, 80%, 90%, 95% or more identity with the residues 51-778 of SEQ. ID No. 1; and performing rational drug design using a three dimensional structure that is based on the obtained structure coordinates. The protein crystals may optionally have a crystal lattice having unit cell dimensions, +/- 5%, of a=121.53Å b=124.11Å and c=144.42Å, $\alpha=\gamma=90^{\circ}$, $\beta=114.6^{\circ}$. The method may optionally further comprise selecting one or more entities based on the rational drug design and contacting the selected entities with the protein. The method may also optionally further comprise measuring an activity of the protein when contacted with the one or more entities. The method also may optionally further comprise comparing activity of the protein in a presence of and in the absence of the one or more entities; and selecting entities where activity of the protein changes depending whether a particular entity is present. The method also may optionally further comprise contacting cells expressing the protein with the one or more entities and detecting a change in a phenotype of the cells when a particular entity is present.

BRIEF DESCRIPTION OF THE FIGURES

[0036] Figure 1 illustrates SEQ. ID Nos. 1, 2 and 3 referred to in this application.

[0037] Figure 2 illustrates a crystal of DPPIV complex.

[0038] Figure 3 lists a set of atomic structure coordinates for DPPIV as derived by X-ray crystallography from a crystal that comprises the protein. The following abbreviations are used in Figure 3: "X, Y, Z" crystallographically define the atomic position of the element measured; "B" is a thermal factor that measures movement of the atom around its atomic center; "Occ" is an occupancy factor that refers to the fraction of the molecules in which each atom occupies the position specified by the coordinates (a value of "1" indicates that each atom has the same conformation, i.e., the same position, in all molecules of the crystal).

[0039] Figure 4A illustrates a ribbon diagram overview of the structure of DPPIV, highlighting secondary structural elements of the protein.

[0040] Figure 4B illustrates another ribbon diagram overview of the structure of DPPIV, highlighting additional secondary structural elements of the protein.

[0041] Figure 5 illustrates the DPPIV binding site of DPPIV based on the determined crystal structure for the molecule in the asymmetric unit corresponding to the coordinates shown in Figure 3.

[0042] Figure 6 illustrates a system that may be used to carry out instructions for displaying a crystal structure of DPPIV encoded on a storage medium.

DETAILED DESCRIPTION OF THE INVENTION

[0043] The present invention relates to a member of the S9 family of human proteases known as dipeptidyl peptidase IV (DPPIV). More specifically, the present invention relates to DPPIV in crystalline form, methods of forming crystals comprising DPPIV, methods of using crystals comprising DPPIV, structure coordinates and a crystal structure of DPPIV, and methods of using the structure coordinates and crystal structure.

[0044] In describing protein structure and function herein, reference is made to amino acids comprising the protein. The amino acids may also be referred to by their conventional abbreviations; A = Ala = Alanine; T = Thr = Threonine; V = Val = Valine; C = Cys = Cysteine; L = Leu = Leucine; Y = Tyr = Tyrosine; I = Ile = Isoleucine; N = Asn = Asparagine; P = Pro = Proline; Q = Gln = Glutamine; F = Phe = Phenylalanine; D = Asp = Aspartic Acid; W = Trp = Tryptophan; E = Glu = Glutamic Acid; M = Met = Methionine; K = Lys = Lysine; G = Gly = Glycine; R = Arg = Arginine; S = Ser = Serine; and H = His = Histidine.

1. DPPIV

[0045] Dipeptidyl Peptidase IV (DPPIV) is a serine protease of Clan SC family S9. DPPIV is a 240kDa homodimeric, multi-functional type-II membrane bound glycoprotein, widely distributed in all mammalian tissues, but highly expressed in kidney, liver and endothelium. DPPIV is also known as DPP4, CD26, adenosine deaminase complexing protein 2 or adenosine deaminase binding protein (ADAbp). DPPIV consists of a short cytoplasmic

domain of six amino acids, followed by a hydrophobic transmembrane domain (amino acids 7-28) and an extracellular sequence of 739 amino acids.

[0046] DPPIV is a highly specific aminopeptidase and releases dipeptides from the amino terminus of peptides with a Pro or Ala in the penultimate position. N-terminal degradation of the substrate peptides may result in the activation, inactivation or modulation of their activity. Besides its well-known exopeptidase activity, DPPIV also exhibits endopeptidase activity towards denatured collagen. Expression of DPPIV is tightly associated with cell adhesion and is a co-stimulant during T-cell activation and proliferation.

[0047] The nature of its substrates, together with its regulated expression and non-enzymatic interactions characterize active participation of DPPIV in the immune, nerve and endocrine networks in human physiology. Among the substrates of DPPIV are glucagon-like peptide 1 (GLP-1) and glucose-dependent insulinotropic polypeptide (GIP), two hormones important for glucose regulation. Degradation and concomitantly inactivation of GLP-1 and GIP by DPPIV reduces insulin secretion.

[0048] It should be understood that the methods and compositions provided relating to DPPIV are not intended to be limited to the wild type, full length form of DPPIV. Instead, the present invention also relates to fragments and variants of DPPIV as described herein. Further, the present invention has applicability to other S9 proteases whose sequence and/or structure are comparatively similar to DPPIV.

[0049] In one embodiment, DPPIV comprises the wild-type form of full length DPPIV, set forth herein as SEQ. ID No. 1 (GenBank Accession Number NM_001935; "Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer cell lines HT-29 and Caco-2. Cloning of the complete human coding sequence and changes of dipeptidyl peptidase IV mRNA levels during cell differentiation", Darmoul, D., Lacasa, M., Baricault, L., Marguet, D., Sapin, C., Trotot, P., Barbat, A. and Trugnan, G., *J. Biol. Chem.* 267 (7), 4824-4833, 1992.

[0050] In another embodiment, DPPIV comprises residues 51-778 of SEQ. ID No. 1 which comprises the active site domain of wild-type DPPIV that is represented in the set of structural coordinates shown in Figure 3.

[0051] It should be recognized that the invention may be readily extended to various variants of wild-type DPPIV and variants of fragments thereof. In another embodiment,

DPPIV comprises a sequence that has at least 65% identity, preferably at least 70%, 80%, 90%, 95% or higher identity with SEQ. ID No. 1.

[0052] It is also noted that the above sequences of DPPIV are also intended to encompass isoforms, mutants and fusion proteins of these sequences. An example of a fusion protein is provided by SEQ. ID No. 3, which includes a 12 residue N-terminal tag (6 residues of which are histidine) that may be used to facilitate purification of the protein.

[0053] With the crystal structure provided herein, where amino acid residues are positioned in the structure are now known. As a result, the impact of different substitutions can be more easily predicted and understood.

[0054] For example, based on the crystal structure, applicants have determined that the DPPIV amino acids shown in Table 1 encompass a 4-Angstrom radius around the DPPIV active site and thus likely to interact with any active site inhibitor of DPPIV. Applicants have also determined that the amino acids of Table 2 encompass a 7-Angstrom radius around the DPPIV active site. Further it has been determined that the amino acids of Table 3 encompass a 10-Angstrom radius around the DPPIV active site. It is noted that there are four different DPPIV molecules in the asymmetric unit, referred to as chains A, B, C and D. As a result, four sets of structure coordinates were obtained for each amino acid. There are two dimers formed in the asymmetric unit; one dimer is formed between molecules A and B and the other with molecules C and D. Applicants have also determined that amino acids of Table 4 encompass a 5-Angstrom radius around the DPPIV amino acids that interact at the AB and CD dimerization interfaces. The A, B, C and D sets of structural coordinates appear in Figure 3. It is noted that the sequence and structure of the residues in the active site and dimerization interface may also be conserved and hence pertinent to other S9 proteases.

[0055] One or more of the sets of amino acids set forth in the tables is preferably conserved in a variant of DPPIV. Hence, DPPIV may optionally comprise a sequence that has at least 65% identity, preferably at least 70%, 80%, 90%, 95% or higher identity with any one of the above sequences (e.g., all of SEQ. ID No. 1 or residues 51-778 of SEQ. ID No. 1) where at least the residues shown in Tables 1, 2, 3 and/or 4 are conserved with the exception of 0, 1, 2, 3, or 4 residues. It should be recognized that one might optionally vary some of the binding site residues in order to determine the effect such changes have on structure or activity.

Table 1: Amino Acids encompassed by a 4-Angstrom radius around the DPPIV active site.

ARG 137	TYR 559	TYR 678
GLU 217	SER 642	ASN 722
GLU 218	TYR 643	HIS 752
SER 221	VAL 668	ASP 720
PHE 369	TYR 674	

Table 2: Amino Acids encompassed by a 7-Angstrom radius around the DPPIV active site.

ARG 137	TYR 559	TRP 671
HIS 138	GLY 561	TYR 674
TRP 213	PRO 562	ASP 675
GLU 216	TYR 597	TYR 678
GLU 217	TRP 641	THR 679
GLU 218	SER 642	ARG 681
VAL 219	TYR 643	TYR 682
PHE 220	GLY 644	ASN 722
SER 221	TYR 646	VAL 723
ARG 368	ALA 666	HIS 752
PHE 369	PRO 667	ASP 720
ARG 370	VAL 668	

Table 3: Amino Acids encompassed by a 10-Angstrom radius around the DPPIV active site.

ARG 137	ILE 417	SER 669
HIS 138	VAL 558	ARG 670
TRP 213	TYR 559	TRP 671
VAL 214	ALA 560	TYR 673
TYR 215	GLY 561	TYR 674
GLU 216	PRO 562	ASP 675
GLU 217	CYS 563	SER 676
GLU 218	SER 564	VAL 677
VAL 219	TYR 597	TYR 678
PHE 220	MET 603	THR 679
SER 221	LEU 610	GLU 680
ALA 222	GLU 614	ARG 681
TYR 268	GLY 640	TYR 682
CYS 313	TRP 641	MET 683
GLN 332	SER 642	HIS 716

TRP 365	TYR 643	ASP 720
VAL 366	GLY 644	ASP 721
GLY 367	GLY 645	ASN 722
ARG 368	TYR 646	VAL 723
PHE 369	VAL 647	HIS 724
ARG 370	VAL 665	GLN 727
PRO 371	ALA 666	HIS 752
SER 372	PRO 667	GLY 753
GLU 373	VAL 668	

Table 4: Amino Acids encompassed by a 5-Angstrom radius around the AB and CD dimerization interfaces.

Chain A	Chain B	Chain C	Chain D
SER A 251	PRO B 246	PRO C 246	LEU D 247
TYR A 253	ILE B 248	LEU C 247	ILE D 248
SER A 254	GLU B 249	ILE C 248	GLU D 249
ASP A 255	TYR B 250	GLU C 249	TYR D 250
GLU A 256	SER B 251	TYR C 250	SER D 251
LEU A 258	TYR B 253	TYR C 253	SER D 254
GLN A 259	SER B 254	SER C 254	ASP D 255
TYR A 260	ASP B 255	ASP C 255	GLU D 256
PRO A 261	GLU B 256	GLU C 256	SER D 257
LYS A 262	SER B 257	SER C 257	LEU D 258
THR A 263	TYR B 260	LEU C 258	GLN D 259
ARG A 265	THR B 263	GLN C 259	TYR D 260
TYR A 268	ARG B 265	TYR C 260	PRO D 261
LYS A 270	GLN B 726	PRO C 261	THR D 263
ALA A 271	ALA B 729	THR C 263	ARG D 265
SER A 732	GLN B 730	TYR C 268	LYS D 270
LYS A 733	LYS B 733	PRO C 269	ALA D 271
LEU A 735	LEU B 735	LYS C 270	ALA D 273
VAL A 736	VAL B 736	GLU C 672	TYR D 673
ASP A 737	ASP B 737	THR C 699	MET D 701
GLY A 739	GLY B 739	LEU C 714	HIS D 716
VAL A 740	VAL B 740	PHE C 725	GLN D 726
PHE A 742	ASP B 741	GLN C 726	SER D 728
GLN A 743	PHE B 742	SER C 728	GLN D 730
MET A 745	GLN B 743	GLN C 730	SER D 732
TRP A 746	ALA B 744	LEU C 735	ASP D 737
TYR A 747		VAL C 736	VAL D 740

THR A 748		ASP C 737	ASP D 741
ASP A 749		GLY C 739	PHE D 742
		VAL C 740	GLN D 743
		ASP C 741	ALA D 744
		PHE C 742	MET D 745
		GLN C 743	TRP D 746
		ALA C 744	TYR D 747
		MET C 745	THR D 748
		TRP C 746	
		TYR C 747	

[0056] With the benefit of the crystal structure and guidance provided by Tables 1, 2, 3 and 4, a wide variety of DPPIV variants (e.g., insertions, deletions, substitutions, etc.) that fall within the above specified identity ranges may be designed and manufactured utilizing recombinant DNA techniques well known to those skilled in the art, particularly in view of the knowledge of the crystal structure provided herein. These modifications can be used in a number of combinations to produce the variants. The present invention is useful for crystallizing and then solving the structure of the range of variants of DPPIV.

[0057] Variants of DPPIV may be insertional variants in which one or more amino acid residues are introduced into a predetermined site in the DPPIV sequence. For instance, insertional variants can be fusions of heterologous proteins or polypeptides to the amino or carboxyl terminus of the subunits.

[0058] Variants of DPPIV also may be substitutional variants in which at least one residue has been removed and a different residue inserted in its place. Non-natural amino acids (i.e. amino acids not normally found in native proteins), as well as isosteric analogs (amino acid or otherwise) may optionally be employed in substitutional variants. Examples of suitable substitutions are well known in the art, such as the Glu→Asp, Ser→Cys, Cys→Ser, and His→Ala for example.

[0059] Another class of variants is deletional variants, which are characterized by the removal of one or more amino acid residues from the DPPIV sequence.

[0060] Other variants may be produced by chemically modifying amino acids of the native protein (e.g., diethylpyrocarbonate treatment that modifies histidine residues). Preferred

are chemical modifications that are specific for certain amino acid side chains. Specificity may also be achieved by blocking other side chains with antibodies directed to the side chains to be protected. Chemical modification includes such reactions as oxidation, reduction, amidation, deamidation, or substitution of bulky groups such as polysaccharides or polyethylene glycol.

[0061] Exemplary modifications include the modification of lysinyl and amino terminal residues by reaction with succinic or other carboxylic acid anhydrides. Modification with these agents has the effect of reversing the charge of the lysinyl residues. Other suitable reagents for modifying amino-containing residues include imidoesters such as methyl picolinimidate; pyridoxal phosphate; pyridoxal chloroborohydride; trinitrobenzenesulfonic acid; 0-methylisourea, 2,4-pentanedione; and transaminaseN: talyzed reaction with glyoxylate, and N-hydroxysuccinamide esters of polyethylene glycol or other bulky substitutions.

[0062] Arginyl residues may be modified by reaction with a number of reagents, including phenylglyoxal, 2,3-butanedione, 1,2-cyclohexanedione, and ninhydrin. Modification of arginine residues requires that the reaction be performed in alkaline conditions because of the high pK_a, of the guanidine functional group. Furthermore, these reagents may react with the groups of lysine as well as the arginine epsilon-amino group.

[0063] Tyrosyl residues may also be modified to introduce spectral labels into tyrosyl residues by reaction with aromatic diazonium compounds or tetranitromethane, forming 0-acetyl tyrosyl species and 3-nitro derivatives, respectively. Tyrosyl residues may also be iodinated using ^{125}I or ^{131}I to prepare labeled proteins for use in radioimmunoassays.

[0064] Carboxyl side groups (aspartyl or glutamyl) may be selectively modified by reaction with carbodiimides or they may be converted to asparaginyl and glutaminyl residues by reaction with ammonium ions. Conversely, asparaginyl and glutaminyl residues may be deamidated to the corresponding aspartyl or glutamyl residues, respectively, under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

[0065] Other modifications that may be formed include the hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl groups of lysine, arginine and histidine side chains (T. E. Creighton, *Proteins: Structure and Molecular Properties*, W.H. Freeman & Co., San Francisco, pp. 79-86, 1983), acetylation of the N-terminal amine and amidation of any C-terminal carboxyl group.

[0066] As can be seen, modifications of the nucleic sequence encoding DPPIV may be accomplished by a variety of well-known techniques, such as site-directed mutagenesis (see, Gillman and Smith, *Gene* 8:81-97 (1979) and Roberts, S. *et al.*, *Nature* 328:731-734 (1987)). When modifications are made, these modifications may optionally be evaluated for there affect on a variety of different properties including, for example, solubility, crystallizability and a modification to the protein's structure and activity.

[0067] In one variation, the variant and/or fragment of wild-type DPPIV is functional in the sense that the resulting protein is capable of associating with at least one same chemical entity that is also capable of selectively associating with a protein comprising the wild-type DPPIV (e.g., residues 51-778 of SEQ. ID No. 1) since this common associative ability evidences that at least a portion of the native structure has been conserved. That chemical entity may optionally be glucagon-like peptide 1 (GLP-1), glucagon-like peptide 2 (GLP-2), glucose-dependent, insulinotropic polypeptide (GIP), growth hormone releasing factor, SDF-1α, β-Casomorphin, TNF-α, Peptide YY or Substance P.

[0068] It is noted the activity of the native protein need not necessarily be conserved. Rather, amino acid substitutions, additions or deletions that interfere with native activity but which do not significantly alter the three-dimensional structure of the domain are specifically contemplated by the invention. Crystals comprising such variants of DPPIV, and the atomic structure coordinates obtained there from, can be used to identify compounds that bind to the native domain. These compounds may affect the activity of the native domain.

[0069] Amino acid substitutions, deletions and additions that do not significantly interfere with the three-dimensional structure of DPPIV will depend, in part, on the region where the substitution, addition or deletion occurs in the crystal structure. These modifications to the protein can now be made far more intelligently with the crystal structure information provided herein. In highly variable regions of the molecule, non-conservative substitutions as well as conservative substitutions may be tolerated without significantly disrupting the three-dimensional structure of the molecule. In highly conserved regions, or regions containing significant secondary structure, conservative amino acid substitutions are preferred.

[0070] Conservative amino acid substitutions are well known in the art, and include substitutions made on the basis of similarity in polarity, charge, solubility, hydrophobicity,

hydrophilicity and/or the amphipathic nature of the amino acid residues involved. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; amino acids with uncharged polar head groups having similar hydrophilicity values include the following: leucine, isoleucine, valine; glycine, alanine; asparagine, glutamine; serine, threonine; phenylalanine, tyrosine. Other conservative amino acid substitutions are well known in the art.

[0071] It should be understood that the protein may be produced in whole or in part by chemical synthesis. As a result, the selection of amino acids available for substitution or addition is not limited to the genetically encoded amino acids. Indeed, mutants may optionally contain non-genetically encoded amino acids. Conservative amino acid substitutions for many of the commonly known non-genetically encoded amino acids are well known in the art. Conservative substitutions for other amino acids can be determined based on their physical properties as compared to the properties of the genetically encoded amino acids.

[0072] In some instances, it may be particularly advantageous or convenient to substitute, delete and/or add amino acid residues in order to provide convenient cloning sites in cDNA encoding the polypeptide, to aid in purification of the polypeptide, etc. Such substitutions, deletions and/or additions which do not substantially alter the three dimensional structure of DPPIV will be apparent to those having skills in the art, particularly in view of the three dimensional structure of DPPIV provided herein.

2. Cloning, Expression and Purification of DPPIV

[0073] The gene encoding DPPIV can be isolated from RNA, cDNA or cDNA libraries. In this case, the portion of the gene encoding amino acid residues 51-778 (SEQ. ID No. 1), corresponding to the catalytic domain of human DPPIV, was isolated and is shown as SEQ. ID No. 2.

[0074] Construction of expression vectors and recombinant proteins from the DNA sequence encoding DPPIV may be performed by various methods well known in the art. For example, these techniques may be performed according to Sambrook et al., Molecular Cloning-A Laboratory Manual, Cold Spring Harbor, N.Y. (1989), and Kriegler, M., Gene Transfer and Expression, A Laboratory Manual, Stockton Press, New York (1990).

[0075] A variety of expression systems and hosts may be used for the expression of DPPIV. Example 1 provides one such expression system.

[0076] Once expressed, purification steps are employed to produce DPPIV in a relatively homogeneous state. In general, a higher purity solution of a protein increases the likelihood that the protein will crystallize. Typical purification methods include the use of centrifugation, partial fractionation, using salt or organic compounds, dialysis, conventional column chromatography, (such as ion exchange, molecular sizing chromatography, etc.), high performance liquid chromatography (HPLC), and gel electrophoresis methods (see, e.g., Deutcher, "Guide to Protein Purification" in Methods in Enzymology (1990), Academic Press, Berkeley, California).

[0077] DPPIV may optionally be affinity labeled during cloning, preferably with an N-terminal six-histidine tag, in order to facilitate purification. With the use of an affinity label, it is possible to perform a one-step purification process on a purification column that has a unique affinity for the label. The affinity label may be optionally removed after purification. These and other purification methods are known and will be apparent to one of skill in the art.

3. Crystallization and Crystals Comprising DPPIV

[0078] One aspect of the present invention relates to methods for forming crystals comprising DPPIV as well as crystals comprising DPPIV.

[0079] In one embodiment, a method for forming crystals comprising DPPIV is provided comprising forming a crystallization volume comprising DPPIV, one or more precipitants, optionally a buffer, optionally a monovalent and/or divalent salt and optionally an organic solvent; and storing the crystallization volume under conditions suitable for crystal formation.

[0080] In yet another embodiment, a method for forming crystals comprising DPPIV is provided comprising forming a crystallization volume comprising DPPIV in solution comprising the components shown in Table 5; and storing the crystallization volume under conditions suitable for crystal formation.

Table 5

Precipitant 5-50% w/v of precipitant wherein the precipitant comprises one or more members of the group consisting of PEG MME having a molecular weight range between 300-10000, and PEG having a molecular weight range between 100-10000 pH pH 5-9. Buffers that may be used include, but are not limited to tris, bicine, cacodylate, acetate, citrate, MES and combinations thereof. Additives optionally 0.05 to 0.8M additives wherein the additives comprises sarcosine or 0.5 to 25% additives wherein the additives comprises xylitrol Protein Concentration 1 mg/ml - 50 mg/ml Temperature 1°C - 25°C

[0081] In yet another embodiment, a method for forming crystals comprising DPPIV is provided comprising forming a crystallization volume comprising DPPIV; introducing crystals comprising DPPIV as nucleation sites, and storing the crystallization volume under conditions suitable for crystal formation.

[0082] Crystallization experiments may optionally be performed in volumes commonly used in the art, for example typically 15, 10, 5, 2 microliters or less. It is noted that the crystallization volume optionally has a volume of less than 1 microliter, optionally 500, 250, 150, 100, 50 or less nanoliters.

[0083] It is also noted that crystallization may be performed by any crystallization method including, but not limited to batch, dialysis and vapor diffusion (e.g., sitting drop and hanging drop) methods. Micro and/or macro seeding of crystals may also be performed to facilitate crystallization.

[0084] It should be understood that forming crystals comprising DPPIV and crystals comprising DPPIV according to the invention are not intended to be limited to the wild type, full length DPPIV shown in SEQ. ID No. 1 and to fragments comprising residues 51-778 of SEQ. ID No. 1. Rather, it should be recognized that the invention may be extended to various other fragments and variants of wild-type DPPIV as described above.

[0085] It should also be understood that forming crystals comprising DPPIV and crystals comprising DPPIV according to the invention may be such that DPPIV is optionally complexed with one or more ligands and one or more copies of the same ligand. The ligand used to form the complex may be any ligand capable of binding to DPPIV. In one variation, the ligand is a natural substrate. In another variation, the ligand is an inhibitor.

[0086] In one particular embodiment, DPPIV crystals have a crystal lattice in the $P2_1$ space group. DPPIV crystals may also optionally have unit cell dimensions, +/- 5%, of a=121.53Å b=124.11Å and c=144.42Å, $\alpha=\gamma=90^{\circ}$, $\beta=114.6^{\circ}$. DPPIV crystals also preferably are capable of diffracting X-rays for determination of atomic coordinates to a resolution of 4 Å, 3 Å, 2.5 Å, 2 Å or better.

[0087] Crystals comprising DPPIV may be formed by a variety of different methods known in the art. For example, crystallizations may be performed by batch, dialysis, and vapor diffusion (sitting drop and hanging drop) methods. A detailed description of basic protein crystallization setups may be found in McRee, D. and David. P., Practical Protein Crystallography, 2nd Ed. (1999), Academic Press Inc. Further descriptions regarding performing crystallization experiments are provided in Stevens, et al. (2000) *Curr. Opin. Struct. Biol.*: 10(5):558-63, and U.S. Patent Nos. 6,296,673, 5,419,278, and 5,096, 676.

[0088] In one variation, crystals comprising DPPIV are formed by mixing substantially pure DPPIV with an aqueous buffer containing a precipitant at a concentration just below a concentration necessary to precipitate the protein. One suitable precipitant for crystallizing DPPIV is polyethylene glycol (PEG), which combines some of the characteristics of the salts and other organic precipitants (see, for example, Ward et al., J. Mol. Biol. 98:161, 1975, and McPherson, J. Biol. Chem. 251:6300, 1976.

[0089] During a crystallization experiment, water is removed by diffusion or evaporation to increase the concentration of the precipitant, thus creating precipitating

conditions for the protein. In one particular variation, crystals are grown by vapor diffusion in hanging drops or sitting drops. According to these methods, a protein/precipitant solution is formed and then allowed to equilibrate in a closed container with a larger aqueous reservoir having a precipitant concentration for producing crystals. The protein/precipitant solution continues to equilibrate until crystals grow.

[0090] By performing submicroliter volume sized crystallization experiments, as detailed in U.S. Patent No. 6,296,673, effective crystallization conditions for forming crystals of a DPPIV complex were obtained. In order to accomplish this, systematic broad screen crystallization trials were performed on a DPPIV complex using the sitting drop technique. Over 1000 individual trials were performed in which pH, temperature and precipitants were varied. In each experiment, a 100nL mixture of DPPIV complex and precipitant was placed on a platform positioned over a well containing 100μL of the precipitating solution. Precipitate and crystal formation was detected in the sitting drops. Fine screening was then carried out for those crystallization conditions that appeared to produce precipitate and/or crystal in the drops.

[0091] Based on the crystallization experiments that were performed, a thorough understanding of how different crystallization conditions affect DPPIV crystallization was obtained. Based on this understanding, a series of crystallization conditions were identified that may be used to form crystals comprising DPPIV. These conditions are summarized in Table 5. A particular example of crystallization conditions that may be used to form diffraction quality crystals of the DPPIV complex is detailed in Example 2. Figure 2 illustrates crystals of the DPPIV complex formed using the crystallization conditions provided in Table 5.

[0092] One skilled in the art will recognize that the crystallization conditions provided in Table 5 and Example 2 can be varied and still yield protein crystals comprising DPPIV. For example, it is noted that variations on the crystallization conditions described herein can be readily determined by taking the conditions provided in Table 5 and performing fine screens around those conditions by varying the type and concentration of the components in order to determine additional suitable conditions for crystallizing DPPIV, variants of DPPIV, and ligand complexes thereof.

[0093] Crystals comprising DPPIV have a wide range of uses. For example, now that crystals comprising DPPIV have been produced, it is noted that crystallizations may be

performed using such crystals as a nucleation site within a concentrated protein solution. According to this variation, a concentrated protein solution is prepared and crystalline material (microcrystals) are used to 'seed' the protein solution to assist nucleation for crystal growth. If the concentrations of the protein and any precipitants are optimal for crystal growth, the seed crystal will provide a nucleation site around which a larger crystal forms. Given the ability to form crystals comprising DPPIV according to the present invention, the crystals so formed can be used by this crystallization technique to initiate crystal growth of other DPPIV comprising crystals, including DPPIV complexed to other ligands.

[0094] As will be described herein in greater detail, crystals may also be used to perform X-ray or neutron diffraction analysis in order to determine the three-dimensional structure of DPPIV and, in particular, to assist in the identification of its active site. Knowledge of the binding site region allows rational design and construction of ligands including inhibitors. Crystallization and structural determination of DPPIV mutants having altered bioactivity allows the evaluation of whether such changes are caused by general structure deformation or by side chain alterations at the substitution site.

4. X-Ray Data Collection and Structure Determination

[0095] Crystals comprising DPPIV may be obtained as described above in Section 3. As described herein, these crystals may then be used to perform x-ray data collection and for structure determination.

[0096] In one embodiment, described in Example 2, crystals of a DPPIV complex were obtained where DPPIV has the sequence of residues shown in SEQ. ID No. 3. These particular crystals were used to determine the three dimensional structure of DPPIV. However, it is noted that other crystals comprising DPPIV including different DPPIV variants, fragments, and complexes thereof may also be used.

[0097] The structure of DPPIV was solved by a combination of heavy-atom derivatives and Seleno-Methionine (Se-Met) phasing in conjunction with non-crystallographic averaging. Heavy atom derivatives were obtained by soaking native DPPIV crystals in heavy atom solutions made using the crystallization solution. The concentration of heavy atom derivative and time of soaking varied between 0.5mM to 10mM and 1 to 15 days, respectively. An

extensive array of heavy atom derivatives were individually soaked into DPPIV crystals and analyzed. Two heavy atom derivatives were used to determine the phases: di-m-iodobis (ethylenediamine)-di-platinum (II) nitrate (PIP) and ethyl mercuric thiosalicylic acid sodium salt (EMTS). Data from crystals of apo DPPIV, Se-Met-DPPIV, PIP-DPPIV and EMTS-DPPIV were collected from cryocooled crystals (100K) at the Stanford Synchrotron Radiation Laboratory (SSRL) beam lines 9-1, 9-2 and 11-1 and the Advanced Light Source (ALS) beam lines 5.0.2 and 5.0.3 both using an ADSC Quantum CCD detector. The diffraction pattern of the DPPIV crystals displayed symmetry consistent with space group $P2_1$ with unit cell dimensions a=121.53Å b=124.11Å and c=144.42Å, $\alpha=\gamma=90^{o}$, $\beta=114.6^{o}$ (+/- 5%). Data were collected and integrated to 2.3Å with MOSFLM (or HKL2000) and scaled with SCALA (or Scalepack) (CCP4 Study Weekend, Eds. Sawyer, L., Isaacs, N. & Bailey, S. 56-62, SERC Daresbury Laboratory, England, 1993).

[0098] All crystallographic calculations were performed using the CCP4 program package (Collaborative Computational Project, N. The CCP4 Suite: Programs for Protein Crystallography. Acta Cryst. D50, 760-763 (1994)).

[0099] Positions for the Platinum atoms of the PIP derivative were located using the direct method search program, SHELXD. The heavy atom parameters of the PIP derivative were refined using the program SHARP. The refined parameters were used to compute phases and locate the Mercury atom positions of the EMTS derivative. The heavy atom parameters of both derivatives were refined using SHARP. Initial solvent flattened maps using the phases from both heavy atom derivatives were of reasonable quality and helped identifying parts of the secondary structure elements of DPPIV. Due to low incorporation of Selenium (Se) in the baculovirus expressed protein, solving Se positions using MAD data was not successful. However, using the phases from the two heavy atom derivatives and cross phasing on to the peak data of the Se-Met derivative of DPPIV allowed for locating all 52 Se atoms of the four subunits of the Se-Met-DPPIV crystal. A final refinement of both the heavy atom derivatives, including the Se atoms, was carried out using SHARP. The resulting phases with solvent flattening and non-crystallographic averaging using DM resulted in an interpretable electron density map. The model was built into the electron density map using Xfit. Refinement continued with iterative map/model/phase improvement using ARP_WARP map (Perrakis, A.,

Morris, R.J. & Lamzin, V.S.). This was followed by alternating cycles of manual rebuilding of the model with Xfit (McRee, D.E. XtalView/Xfit-A versatile program for manipulating atomic coordinates and electron density J. Struct. Biol. 125, 156-65 (1999)), ARP_WARP map improvement (Perrakis, A., Morris, R.J. & Lamzin, V.S. Automated protein model building combined with iterative structure refinement) and geometrically restrained refinement against a maximum likelihood target function as implemented in REFMAC (CCP4) until the refinement reached convergence. All stages of model refinement were carried with bulk solvent correction and anisotropic scaling. The data collection and data refinement statistics are given in Table 6.

TABLE 6

Crystal data	
Space group	$P2_1$
Unit cell dimensions	a=121.53Å b=124.11Å and c=144.42Å, $\alpha=\gamma=90^\circ$, $\beta=114.6^\circ$
Data collection	
X-ray source	ALS: BL5.0.2; BL5.0.3; SSRL: BL9-1; BL9-2; BL11-1
Wavelength [Å]	0.90 to 1.25
Resolution [Å]	2.30
Observations (unique)	183144
Redundancy	2.8
Completeness overall (outer shell)	98(97.6)%
$I/\sigma(I)$ overall (outer shell)	8.6 (1.9)
R_{symm}[1] overall (outer shell)	8.2 (51.6)%
Refinement	
Reflections used	159715
R-factor	22.7%
R_{free}	28.3%
r.m.s bonds	0.009Å
r.m.s angles	1.432°

[00100] During structure determination, where the unit cell dimensions were a=121.53Å b=124.11Å and c=144.42Å, $\alpha=\gamma=90^{o}$, $\beta=114.6^{o}$, it was realized that each unit cell comprised four DPPIV molecules. Structure coordinates were determined for this complex and the resultant set of structural coordinates from the refinement are presented in Figure 3.

[00101] It is noted that the sequence of the structure coordinates presented in Figure 3 differ in some regards from the sequence shown in SEQ. ID No. 1. Structure coordinates are not reported for some residues because the electron density obtained was insufficient to identify the position of these residues. For Figure 3, structure coordinates for residues 151-153 (chains C and D) and 97-99 (chain D) are not reported.

[00102] Those of skill in the art understand that a set of structure coordinates (such as those in Figure 3) for a protein or a protein-complex or a portion thereof, is a relative set of points that define a shape in three dimensions. Thus, it is possible that an entirely different set of structure coordinates could define a similar or identical shape. Moreover, slight variations in the individual coordinates may have little effect on overall shape. In terms of binding pockets, these variations would not be expected to significantly alter the nature of ligands that could associate with those pockets. The term "binding pocket" as used herein refers to a region of the protein that, as a result of its shape, favorably associates with a ligand.

[00103] These variations in coordinates may be generated because of mathematical manipulations of the DPPIV structure coordinates. For example, the sets of structure coordinates shown in Figure 3 could be manipulated by crystallographic permutations of the structure coordinates, fractionalization of the structure coordinates, application of a rotation matrix, integer additions or subtractions to sets of the structure coordinates, inversion of the structure coordinates or any combination of the above.

[00104] Alternatively, modifications in the crystal structure due to mutations, additions, substitutions, and/or deletions of amino acids or other changes in any of the components that make up the crystal could also account for variations in structure coordinates. If such variations are within an acceptable standard error as compared to the original coordinates, the resulting three-dimensional shape should be considered to be the same. Thus, for example, a ligand that bound to the active site binding pocket of DPPIV would also be expected to bind to

another binding pocket whose structure coordinates defined a shape that fell within the acceptable error.

[00105] Various computational analyses may be used to determine whether structure coordinates for a protein or a portion thereof is similar to the structure coordinates of DPPIV provided herein, or a portion thereof. Such analyses may be carried out in well known software applications, such as the Molecular Similarity application of QUANTA (Molecular Simulations Inc., San Diego, Calif.) version 4.1, and as described in the accompanying User's Guide. For the purpose of this invention, a rigid fitting method shall be used to compare protein structures.

[00106] For the purpose of this invention, any set of structure coordinates for a protein from any source having a root mean square deviation of alpha-carbon atoms of less than 3 Å when superimposed on the alpha-carbon atom positions of the corresponding atomic coordinates of Figure 3 shall be considered identical. It is noted that the root mean square deviation is intended to be limited to only those alpha-carbon atoms of amino acid residues that are common to both the protein fragment represented in Figure 3 and the protein whose structure coordinates are being compared to the coordinates shown in Figure 3.

[00107] It is noted that mutants and variants of DPPIV as well as other S9 proteases are likely to have similar structures despite having different sequences. For example, the binding pockets of these related proteins are likely to have similar contours. Accordingly, it should be recognized that the structure coordinates and binding pocket models provided herein have utility for these other related proteins.

[00108] Accordingly, in one embodiment, the invention relates to data, computer readable media comprising data, and uses of the data where the data comprises all or a portion of the structure coordinates shown in Figure 3 or structure coordinates having a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on the alpha-carbon atom positions of the corresponding atomic coordinates of Figure 3. Again, it is noted that the root mean square deviation is intended to be limited to only those alpha-carbon atoms of amino acid residues that are common to both the protein fragment represented in Figure 3 and the protein whose structure coordinates are being compared to the coordinates shown in Figure 3.

[00109] As noted, there are many different ways to express the surface contours of the DPPIV structure other than by using the structure coordinates provided in Figure 3. Accordingly, it is noted that the present invention is also directed to any data, computer readable media comprising data, and uses of the data where the data defines a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on alpha-carbon atoms in the structure coordinates of Figure 3 that are present in residues shown in SEQ. ID No. 1.

[00110] In regard to these embodiments, it is noted that the root mean square deviation calculation may optionally be based on a comparison of main-chain atoms or non-hydrogen atoms. Also, the root mean square deviation of alpha-carbon atoms, main-chain atoms or non-hydrogen atoms may optionally be less than 2.7 Å, 2.5 Å, 2.0 Å, 1.5 Å, 1 Å, 0.5 Å, or less.

5. DPPIV Structure

[00111] The present invention is also directed to a three-dimensional crystal structure of DPPIV. This crystal structure may be used to identify binding sites, to provide mutants having desirable binding properties, and ultimately, to design, characterize, or identify ligands that interact with DPPIV as well as other S9 proteases.

[00112] The three-dimensional crystal structure of DPPIV may be generated, as is known in the art, from the structure coordinates shown in Figure 3 and similar such coordinates.

[00113] During the course of structure solution it became evident that the wild type apo crystals of DPPIV of the present invention contained four nearly identical copies in the asymmetric unit. The final coordinates for each one of these molecules, referred to as chains A, B, C and D, are given in Figure 3. The variations between the chains are described below.

[00114] Chain A includes amino acid residues 52 to 778 and four amino acid residues have covalently linked sugar molecules (Figure 3). Chain B includes amino acid residues 47 to 778 and also includes 4 histidine residues of the N-terminal polyhistidine tag. Five amino acid residues of chain B have covalently linked sugar molecules (Figure 3). Chain C includes amino

acid residues 52 to 778 and five of the amino acid residues are covalently linked to sugar molecules. Chain D includes amino acid residues 51 to 778 with five sugar-linked amino acid residues. In addition, chains C and D have no density for amino acid residues 151, 152 and 153 and hence coordinates for these residues are not included in Figure 3. Similarly, coordinates for amino acid residues 97, 98 and 99 of chain D are not included in Figure 3. The coordinate set additionally includes 928 solvent molecules modeled as water.

[00115] Figure 4A illustrates a ribbon diagram overview of the structure of DPPIV, highlighting secondary structural elements of the protein. DPPIV is a cylindrical shaped molecule with an approximate height of 70 Å and a diameter of 60 Å (Fig. 4A). The catalytic triad of DPPIV (Ser642, Asp720 and His752) is illustrated in the center of Figure 4A by a "ball and stick" representation. This triad of amino acids is located in the peptidase domain or catalytic domain of DPPIV. The catalytic domain is covalently linked to the β-propeller domain (Fig. 4A).

[00116] The catalytic domain of DPPIV includes residues 1-67 and 511-778. Since, the structure of the present invention does not contain the first 46 residues (Chain B of Figure 3) it is presumed that the N-terminal residues of the catalytic domain adopt a random structure with a short double turn α-helix formed by residues 56 to 63. The catalytic domain of DPPIV adopts a characteristic α/β hydrolase fold. The core of this domain contains an 8-stranded β-sheet with all strands being parallel except one (Fig. 4A). The β-sheet is significantly twisted and is flanked by three α-helices on one side and five α-helices on the other. The topology of the β-strands is 1, 2, -1x, 2x and (1x) (J. S. Richardson: The anatomy and taxonomy of protein structure; (1981) *Adv. Protein Chem.* **269**, 15076-15084.).

[00117] Figure 4B illustrates the remaining residues 68-510 that form the non-catalytic domain of DPPIV. This domain is also known as β-propeller domain (Figure 4B). The β-propeller domain is a 7-fold repeat of four-stranded antiparallel β-sheets (Figure 4B). The sheets are twisted and arranged around a central tunnel as seen in case of Prolyl Oligopeptidase. Further, the β-sheets pack face-to-face and are stabilized predominantly by hydrophobic interactions. The β-propeller is linked to the catalytic domain by two polypeptide chains, one

involving the N-terminal residues and the other consisting of the C-terminal residues 511-520 which also form an α-helix.

[00118] Figure 5 illustrates the binding site of DPPIV based on the determined crystal structure corresponding to the coordinates shown in Figure 3.

6. DPPIV Active Site and Ligand Interaction

[00119] The term "binding site" or "binding pocket", as the terms are used herein, refers to a region of a protein that, as a result of its shape, favorably associates with a ligand or substrate. The term "DPPIV-like binding pocket" refers to a portion of a molecule or molecular complex whose shape is sufficiently similar to the DPPIV binding pockets as to bind common ligands. This commonality of shape may be quantitatively defined based on a comparison to a reference point, that reference point being the structure coordinates provided herein. For example the commonality of shape may be quantitatively defined based on a root mean square deviation (rmsd) from the structure coordinates of the backbone atoms of the amino acids that make up the binding pockets in DPPIV (as set forth in Figure 3).

[00120] The "active site binding pockets" or "active site" of DPPIV refers to the area on the surface of DPPIV where the substrate binds.

[00121] Figure 5 illustrates the inhibitor-binding site of DPPIV based on the determined crystal structure (coordinates shown in Figure 3). The active site containing the catalytic triad (Ser642, Asp 720 and His 752), is located in a large cavity (Figure 5) at the interface of the catalytic and the β-propeller domains. Ser 642 is located on a sharp turn that connects an α-helix to a β-strand. The positioning of this active site Serine residue is referred to as a nucleophile elbow and is characteristic of an α/β type hydrolase (D. J. Ollis et al., The α/β hydrolase fold; (1992) *Protein Eng.* **5**, 197-211). In DPPIV, the active-site serine is surrounded by hydrophobic residues, which include the large aromatic residues Trp 641 and Tyr 643. The hydroxyl group of the active site serine is exposed and involved in hydrogen bonding with the imidazole group of the active site His 752 (OH------NH distance 2.7 Å). His 752 is located on the middle of a loop that connects a β-strand to an α-helix. The other nitrogen atom of the imidazole ring of His 752 forms a hydrogen bond with the side chain oxygen of the third active site residue (Asp 720) of the catalytic triad. Asp 720 is also located

on a loop connecting a β-strand and an α-helix. The second oxygen atom of the side chain carboxylate of Asp 720 forms two hydrogen bonded interactions with the main-chain amide of residues (Asn 722 and Val 723). The hydrogen bonding interactions of the catalytic triad is similar to those observed for prolyl oligopeptidase.

[00122] Based on sequence alignments and structural comparisons with prolyl oligopeptidase, the residues that form the DPPIV active site pocket can be predicted with a high degree of probability. The binding pocket appears to be formed by a pocket of hydrophobic residues (Phe 369, Tyr 643, Tyr 674, Tyr 678, Tyr 559 and Val 723). In addition to the catalytic triad a large number of polar residues are also present in this hydrophobic environment (Arg 137, Glu 217, Glu 218 and Asp 675).

[00123] In resolving the crystal structure of DPPIV, applicants determined that DPPIV amino acids shown in Table 1 (above) are encompassed within a 4-Angstrom radius around the DPPIV active site and therefore are likely close enough to interact with an active site inhibitor of DPPIV. Applicants have also determined that the amino acids shown in Table 2 (above) are encompassed within a 7-Angstrom radius around the DPPIV active site. Further, the amino acids shown in Table 3 (above) are encompassed within a 10-Angstrom radius around the DPPIV active site. Due to their proximity to the active site, the amino acids in the 4, 7, and/or 10 Angstroms sets are preferably conserved in variants of DPPIV. While it is desirable to largely conserve these residues, it should be recognized however that variants may also involve varying 1, 2, 3, 4 or more of the residues set forth in Tables 1, 2, and 3 in order to evaluate the roles these amino acids play in the binding pocket. Applicants have also determined that amino acids shown in Table 4 (above) are encompassed within a 5-Angstrom radius around the DPPIV dimerization interface (AB and CD dimers).

[00124] With the knowledge of the DPPIV crystal structure provided herein, Applicants are able to know the contour of a DPPIV binding pocket as a binding pocket where the relative positioning of the 4, 7, and/or 10 Angstroms sets of amino acids. In addition, Applicants are able to know the contour of a dimerization interface (AB and CD dimers) based on the relative positions of the α-carbon residues in Table 4. Again, it is noted that it may be desirable to form variants where 1, 2, 3, 4 or more of the residues set forth in Tables 1, 2, and 3 are varied in order to evaluate the roles these amino acids play in the binding pocket. Accordingly, any set

of structure coordinates for a protein from any source having a root mean square deviation of non-hydrogen atoms of less than 3 Å when superimposed on the non-hydrogen atom positions of the corresponding atomic coordinates of Figure 3 for the 4, 7, and/or 10 Angstroms sets of amino acids and/or those amino acids of the dimerization interface shall be considered identical. As noted previously, the root mean square deviation is intended to be limited to only those non-hydrogen atoms of amino acid residues that are common to both the protein fragment represented in Figure 3 and the protein whose structure coordinates are being compared to the coordinates shown in Figure 3 since the sequence of the protein may be varied somewhat.

[00125] Accordingly, in one embodiment, the invention relates to data, computer readable media comprising data, and uses of the data where the data comprises the structure coordinates shown in Figure 3 or structure coordinates having a root mean square deviation of non-hydrogen atoms of less than 3Å when superimposed on the non-hydrogen atom positions of the corresponding atomic coordinates of Figure 3 for the 4, 7, and/or 10 Angstroms sets of amino acids and/or the residues listed in Table 4.

[00126] Again, it is noted that the root mean square deviation is intended to be limited to only those non-hydrogen atoms of amino acid residues that are common to both the protein fragment represented in one or more of the tables and the protein whose structure coordinates are being compared to the coordinates shown in Figure 3.

[00127] As noted above, there are many different ways to express the surface contours of the DPPIV structure other than by using the structure coordinates provided in Figure 3. Accordingly, it is noted that the present invention is also directed to any data, computer readable media comprising data, and uses of the data where the data defines a computer model for a protein binding pocket, at least a portion of the computer model having a surface contour that has a root mean square deviation of less than 3Å when superimposed on a surface contour defined by atomic coordinates of Figure 3, the root mean square deviation being calculated based only on non-hydrogen atoms in the structure coordinates of Figure 3 that are present in residues shown in Tables 1, 2, 3 and/or 4.

[00128] Optionally, the root mean square deviation of non-hydrogen atoms is less than 1.5 Å, 1 Å, 0.5 Å, or less.

[00129] It will be readily apparent to those of skill in the art that the numbering of amino acids in other isoforms of DPPIV may be different than that set forth for DPPIV. Corresponding amino acids in other isoforms of DPPIV are easily identified by visual inspection of the amino acid sequences or by using commercially available homology software programs, as further described below.

7. System For Displaying the Three Dimensional Structure of DPPIV

[00130] The present invention is also directed to machine-readable data storage media having data storage material encoded with machine-readable data that comprises structure coordinates for DPPIV. The present invention is also directed to a machine readable data storage media having data storage material encoded with machine readable data, which, when read by an appropriate machine, can display a three dimensional representation of a structure of DPPIV.

[00131] All or a portion of the DPPIV coordinate data shown in Figure 3, when used in conjunction with a computer programmed with software to translate those coordinates into the three-dimensional structure of DPPIV may be used for a variety of purposes, especially for purposes relating to drug discovery. Softwares for generating three-dimensional graphical representations are known and commercially available. The ready use of the coordinate data requires that it be stored in a computer-readable format. Thus, in accordance with the present invention, data capable of being displayed as the three-dimensional structure of DPPIV and/or portions thereof and/or their structurally similar variants may be stored in a machine-readable storage medium, which is capable of displaying a graphical three-dimensional representation of the structure.

[00132] For example, in one embodiment, a computer is provided for producing a three-dimensional representation of at least an DPPIV-like binding pocket, the computer comprising: machine readable data storage medium comprising a data storage material encoded with machine-readable data, the machine readable data comprising structure coordinates that have a root mean square deviation of less than 3 Angstroms when compared to structure coordinates appearing in Figure 3, the comparison being based on alpha-carbon atoms of amino acid residues present in both the set of structure coordinates shown in Figure 3 and the structure coordinates being compared, the comparison being further limited to residues of DPPIV

appearing in Tables 1, 2, 3 and/or 4; a working memory for storing instructions for processing the machine-readable data; a central-processing unit coupled to the working memory and to the machine-readable data storage medium, for processing the machine-readable data into the three-dimensional representation; and an output hardware coupled to the central processing unit, for receiving the three Dimensional representation.

[00133] Another embodiment of this invention provides a machine-readable data storage medium, comprising a data storage material encoded with machine readable data which, when used by a machine programmed with instructions for using said data, displays a graphical three-dimensional representation comprising DPPIV or a portion or variant thereof.

[00134] In one variation, the machine readable data comprises data for representing a protein based on structure coordinates having a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on the alpha-carbon atom positions of the corresponding atomic coordinates of Figure 3 for all of the amino acids in Figure 3.

[00135] In another variation, the machine readable data comprises data for representing a protein based on structure coordinates having a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on the alpha-carbon atom positions of the corresponding atomic coordinates of Figure 3 for the amino acids listed in Tables 1, 2, 3 and/or 4.

[00136] It is again noted that the root mean square deviation calculation may optionally be based on a comparison of main-chain atoms or non-hydrogen atoms. Also, the root mean square deviation of alpha-carbon atoms, main-chain atoms or non-hydrogen atoms may optionally be less than 2.7Å, 2.5Å, 2.0Å, 1.5Å, 1Å, 0.5Å, or less.

[00137] According to another embodiment, the machine-readable data storage medium comprises a data storage material encoded with a first set of machine readable data which comprises the Fourier transform of the structure coordinates set forth in Figure 3, and which, when using a machine programmed with instructions for using said data, can be combined with a second set of machine readable data comprising the X-ray diffraction pattern of another molecule or molecular complex to determine at least a portion of the structure coordinates corresponding to the second set of machine readable data. For example, the Fourier transform

of the structure coordinates set forth in Figure 3 may be used to determine at least a portion of the structure coordinates of other DPPIV-like enzymes, and isoforms of DPPIV.

[00138] Optionally, a computer system is provided in combination with the machine-readable data storage medium provided herein. In one embodiment, the computer system comprises a working memory for storing instructions for processing the machine-readable data; a processing unit coupled to the working memory and to the machine-readable data storage medium, for processing the machine-readable data into the three-dimensional representation; and an output hardware coupled to the processing unit, for receiving the three-dimensional representation.

[00139] Figure 6 illustrates an example of a computer system that may be used in combination with storage media according to the present invention. As illustrated, the computer system 10 includes a computer 11 comprising a central processing unit ("CPU") 20, a working memory 22 which may be, e.g., RAM (random-access memory) or "core" memory, mass storage memory 24 (such as one or more disk drives or CD-ROM drives), one or more cathode-ray tube ("CRT") display terminals 26, one or more keyboards 28, one or more input lines 30, and one or more output lines 40, all of which are interconnected by a conventional bi-directional system bus 50.

[00140] Input hardware 36, coupled to computer 11 by input lines 30, may be implemented in a variety of ways. For example, machine-readable data of this invention may be inputted via the use of a modem or modems 32 connected by a telephone line or dedicated data line 34. Alternatively or additionally, the input hardware 36 may comprise CD-ROM drives or disk drives 24. In conjunction with display terminal 26, keyboard 28 may also be used as an input device.

[00141] Conventional devices may, similarly implement output hardware 46, coupled to computer 11 by output lines 40. By way of example, output hardware 46 may include CRT display terminal 26 for displaying a graphical representation of a binding pocket of this invention using a program such as QUANTA as described herein. Output hardware might also include a printer 42, so that hard copy output may be produced, or a disk drive 24, to store system output for later use.

[00142] In operation, CPU 20 coordinates the use of the various input and output devices 36, 46 coordinates data accesses from mass storage 24 and accesses to and from working memory 22, and determines the sequence of data processing steps. A number of programs may be used to process the machine-readable data of this invention. Such programs are discussed in reference to using the three dimensional structure of DPPIV described herein.

[00143] The storage medium encoded with machine-readable data according to the present invention can be any conventional data storage device known in the art. For example, the storage medium can be a conventional floppy diskette or hard disk. The storage medium can also be an optically-readable data storage medium, such as a CD-ROM or a DVD-ROM, or a rewritable medium such as a magneto-optical disk that is optically readable and magneto-optically writable.

8. Uses of the Three Dimensional Structure of DPPIV

[00144] The three-dimensional crystal structure of the present invention may be used to identify DPPIV binding sites, be used as a molecular replacement model to solve the structure of unknown crystallized proteins, to design mutants having desirable binding properties, and ultimately, to design, characterize, identify entities capable of interacting with DPPIV and other S9 proteases, as well as other uses that would be recognized by one of ordinary skill in the art. Such entities may be chemical entities or proteins. The term "chemical entity", as used herein, refers to chemical compounds, complexes of at least two chemical compounds, and fragments of such compounds.

[00145] The DPPIV structure coordinates provided herein are useful for screening and identifying drugs that inhibit DPPIV and other proteases. For example, the structure encoded by the data may be computationally evaluated for its ability to associate with putative substrates or ligands. Such compounds that associate with DPPIV may inhibit DPPIV, and are potential drug candidates. Additionally or alternatively, the structure encoded by the data may be displayed in a graphical three-dimensional representation on a computer screen. This allows visual inspection of the structure, as well as visual inspection of the structure's association with the compounds.

[00146] Thus, according to another embodiment of the present invention, a method is provided for evaluating the potential of an entity to associate with DPPIV or a fragment or variant thereof by using all or a portion of the structure coordinates provided in Figure 3. A method is also provided for evaluating the potential of an entity to associate with DPPIV or a fragment or variant thereof by using structure coordinates similar to all or a portion of the structure coordinates provided in Figure 3. For example, the structure coordinates used may have a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on the alpha-carbon atom positions of the corresponding atomic coordinates of Figure 3.

[00147] It is again noted in regard to these embodiments that the root mean square deviation calculation may optionally be based on a comparison of main-chain atoms or non-hydrogen atoms. Also, the root mean square deviation of alpha-carbon atoms or non-hydrogen atoms may optionally be less than 2.7Å, 2.5Å, 2.0Å, 1.5Å, 1Å, 0.5Å, or less.

[00148] The method may optionally comprise the steps of: creating a computer model of all or a portion of a protein structure (e.g., a binding pocket) using structure coordinates according to the present invention; performing a fitting operation between the entity and the computer model; and analyzing the results of the fitting operation to quantify the association between the entity and the model. The portion of the protein structure used optionally comprises all of the amino acids listed in Tables 1, 2, 3 and/or 4 that are present in the structure coordinates being used.

[00149] It is noted that the computer model may not necessarily directly use the structure coordinates. Rather, a computer model can be formed that defines a surface contour that is the same or similar to the surface contour defined by the structure coordinates.

[00150] The structure coordinates provided herein can also be utilized in a method for identifying a ligand (e.g., entities capable of associating with a protein) of a protein comprising a DPPIV-like binding pocket. One embodiment of the method comprises: using all or a portion of the structure coordinates provided herein to generate a three-dimensional structure of a DPPIV-like binding pocket; employing the three-dimensional structure to design or select a potential ligand; synthesizing the potential ligand; and contacting the synthesized potential ligand with a protein comprising an DPPIV-like binding pocket to determine the ability of the potential ligand to interact with protein. According to this method, the structure coordinates

used may have a root mean square deviation of alpha-carbon atoms of less than 3Å when superimposed on the alpha-carbon atom positions of the corresponding atomic coordinates of Figure 3. The portion of the protein structure used optionally comprises all of the amino acids listed in Tables 1, 2, 3 and/or 4 that are present.

[00151] As noted previously, the three-dimensional structure of a DPPIV-like binding pocket need not be generated directly from structure coordinates. Rather, a computer model can be formed that defines a surface contour that is the same or similar to the surface contour defined by the structure coordinates.

[00152] It is again noted that the root mean square deviation calculation may optionally be based on a comparison of main-chain or non-hydrogen atoms. Also, the root mean square deviation of alpha-carbon atoms or non-hydrogen atoms may optionally be less than 2.7Å, 2.5Å, 2.0Å, 1.5Å, 1Å, 0.5Å, or less.

[00153] A method is also provided for evaluating the ability of an entity, such as a compound or a protein to associate with a DPPIV-like binding pocket, the method comprising: constructing a computer model of a binding pocket defined by structure coordinates that have a root mean square deviation of less than 3.0Å when compared to structure coordinates appearing in Figure 3, the comparison being based on alpha-carbon atoms present in both sets of structure coordinates, the comparison also being limited to residues of DPPIV appearing in Tables 1, 2, 3 and/or 4 that are present; selecting an entity to be evaluated by a method selected from the group consisting of (i) assembling molecular fragments into the entity, (ii) selecting an entity from a small molecule database, (iii) *de novo* ligand design of the entity, and (iv) modifying a known ligand for DPPIV, or a portion thereof; performing a fitting program operation between computer models of the entity to be evaluated and the binding pocket in order to provide an energy-minimized configuration of the entity in the binding pocket; and evaluating the results of the fitting operation to quantify the association between the entity and the binding pocket model in order to evaluate the ability of the entity to associate with the said binding pocket.

[00154] The computer model of a binding pocket used in this embodiment need not be generated directly from structure coordinates. Rather, a computer model can be formed that defines a surface contour that is the same or similar to the surface contour defined by the structure coordinates.

[00155] According to the method, the root mean square deviation calculation may optionally be based on a comparison of main-chain atoms or non-hydrogen atoms. Also, the root mean square deviation of alpha-carbon atoms or non-hydrogen atoms may optionally be less than 2.7Å, 2.5Å, 2.0Å, 1.5Å, 1Å, 0.5Å, or less.

[00156] Also according to the method, the method may further include synthesizing the entity; and contacting a protein having a DPPIV-like binding pocket with the synthesized entity.

[00157] With the structure provided herein, the present invention for the first time permits the use of molecular design techniques to identify, select or design potential inhibitors of DPPIV, based on the structure of a DPPIV-like binding pocket. Such a predictive model is valuable in light of the high costs associated with the preparation and testing of the many diverse compounds that may possibly bind to the DPPIV protein.

[00158] According to this invention, a potential DPPIV inhibitor may now be evaluated for its ability to bind a DPPIV-like binding pocket prior to its actual synthesis and testing. If a proposed entity is predicted to have insufficient interaction or association with the binding pocket, preparation and testing of the entity can be obviated. However, if the computer modeling indicates a strong interaction, the entity may then be obtained and tested for its ability to bind.

[00159] A potential inhibitor of a DPPIV-like binding pocket may be computationally evaluated using a series of steps in which chemical entities or fragments are screened and selected for their ability to associate with the DPPIV-like binding pockets.

[00160] One skilled in the art may use one of several methods to screen entities (whether chemical or protein) for their ability to associate with a DPPIV-like binding pocket. This process may begin by visual inspection of, for example, a DPPIV-like binding pocket on a computer screen based on the DPPIV structure coordinates in Figure 3 or other coordinates which define a similar shape generated from the machine-readable storage medium. Selected fragments or chemical entities may then be positioned in a variety of orientations, or docked, within that binding pocket as defined above. Docking may be accomplished using software such as Quanta and Sybyl, followed by energy minimization and molecular dynamics with standard molecular mechanics force fields, such as CHARMM and AMBER.

[00161] Specialized computer programs may also assist in the process of selecting entities. These include: GRID (P. J. Goodford, "A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules", J. Med. Chem., 28, pp. 849-857 (1985)). GRID is available from Oxford University, Oxford, UK; MCSS (A. Miranker et al., "Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method." Proteins: Structure, Function and Genetics, 11, pp. 29-34 (1991)). MCSS is available from Molecular Simulations, San Diego, Calif.; AUTODOCK (D. S. Goodsell et al., "Automated Docking of Substrates to Proteins by Simulated Annealing", Proteins: Structure, Function, and Genetics, 8, pp. 195-202 (1990)). AUTODOCK is available from Scripps Research Institute, La Jolla, Calif.; & DOCK (I. D. Kuntz et al., "A Geometric Approach to Macromolecule-Ligand Interactions", J. Mol. Biol., 161, pp. 269-288 (1982)). DOCK is available from University of California, San Francisco, Calif.

[00162] Once suitable entities have been selected, they can be designed or assembled. Assembly may be preceded by visual inspection of the relationship of the fragments to each other on the three-dimensional image displayed on a computer screen in relation to the structure coordinates of DPPIV. This may then be followed by manual model building using software such as Quanta or Sybyl [Tripos Associates, St. Louis, Mo].

[00163] Useful programs to aid one of skill in the art in connecting the individual chemical entities or fragments include: CAVEAT (P. A. Bartlett et al, "CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules", in "Molecular Recognition in Chemical and Biological Problems", Special Pub., Royal Chem. Soc., 78, pp. 182-196 (1989); G. Lauri and P. A. Bartlett, "CAVEAT: a Program to Facilitate the Design of Organic Molecules", J. Comput. Aided Mol. Des., 8, pp. 51-66 (1994)). CAVEAT is available from the University of California, Berkeley, Calif.; 3D Database systems such as ISIS (MDL Information Systems, San Leandro, Calif.). This area is reviewed in Y. C. Martin, "3D Database Searching in Drug Design", J. Med. Chem., 35, pp. 2145-2154 (1992); HOOK (M. B. Eisen et al, "HOOK: A Program for Finding Novel Molecular Architectures that Satisfy the Chemical and Steric Requirements of a Macromolecule Binding Site", Proteins: Struct., Funct., Genet., 19, pp. 199-221 (1994). HOOK is available from Molecular Simulations, San Diego, Calif.

[00164] Instead of proceeding to build an inhibitor of a DPPIV-like binding pocket in a step-wise fashion one fragment or entity at a time as described above, inhibitory or other DPPIV binding compounds may be designed as a whole or "*de novo*" using either an empty binding site or optionally including some portion(s) of a known inhibitor(s). There are many *de novo* ligand design methods including: LUDI (H.-J. Bohm, "The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors", J. Comp. Aid. Molec. Design, 6, pp. 61-78 (1992)). LUDI is available from Molecular Simulations Incorporated, San Diego, Calif.; LEGEND (Y. Nishibata et al., Tetrahedron, 47, p. 8985 (1991)). LEGEND is available from Molecular Simulations Incorporated, San Diego, Calif.; LEAPFROG (available from Tripos Associates, St. Louis, Mo.); & SPROUT (V. Gillet et al, "SPROUT: A Program for Structure Generation)", J. Comput. Aided Mol. Design, 7, pp. 127-153 (1993)). SPROUT is available from the University of Leeds, UK.

[00165] Other molecular modeling techniques may also be employed in accordance with this invention (see, e.g., Cohen et al., "Molecular Modeling Software and Methods for Medicinal Chemistry, J. Med. Chem., 33, pp. 883-894 (1990); see also, M. A. Navia and M. A. Murcko, "The Use of Structural Information in Drug Design", Current Opinions in Structural Biology, 2, pp. 202-210 (1992); L. M. Balbes et al., "A Perspective of Modern Methods in Computer-Aided Drug Design", in Reviews in Computational Chemistry, Vol. 5, K. B. Lipkowitz and D. B. Boyd, Eds., VCH, New York, pp. 337-380 (1994); see also, W. C. Guida, "Software For Structure-Based Drug Design", Curr. Opin. Struct. Biology, 4, pp. 777-781 (1994)).

[00166] Once an entity has been designed or selected, for example, by the above methods, the efficiency with which that entity may bind to a DPPIV binding pocket may be tested and optimized by computational evaluation. For example, an effective DPPIV binding pocket inhibitor preferably demonstrates a relatively small difference in energy between its bound and free states (i.e., a small deformation energy of binding). Thus, the most efficient DPPIV binding pocket inhibitors should preferably be designed with deformation energy of binding of not greater than about 10 kcal/mole, more preferably, not greater than 7 kcal/mole. DPPIV binding pocket inhibitors may interact with the binding pocket in more than one of multiple conformations that are similar in overall binding energy. In those cases, the

deformation energy of binding is taken to be the difference between the energy of the free entity and the average energy of the conformations observed when the inhibitor binds to the protein.

[00167] An entity designed or selected as binding to a DPPIV binding pocket may be further computationally optimized so that in its bound state it would preferably lack repulsive electrostatic interaction with the target enzyme and with the surrounding water molecules. Such non-complementary electrostatic interactions include repulsive charge-charge, dipole-dipole and charge-dipole interactions.

[00168] Specific computer software is available in the art to evaluate compound deformation energy and electrostatic interactions. Examples of programs designed for such uses include: Gaussian 94, revision C (M. J. Frisch, Gaussian, Inc., Pittsburgh, Pa. .COPYRGT.1995); AMBER, version 4.1 (P. A. Kollman, University of California at San Francisco, COPYRGT 1995); QUANTA/CHARMM (Molecular Simulations, Inc., San Diego, Calif. COPYRGT.1995); Insight II/Discover (Molecular Simulations, Inc., San Diego, Calif. COPYRGT.1995); DelPhi (Molecular Simulations, Inc., San Diego, Calif. COPYRGT.1995); and AMSOL (Quantum Chemistry Program Exchange, Indiana University). These programs may be implemented, for instance, using a Silicon Graphics workstation such as an Indigo.sup.2 with "IMPACT" graphics. Other hardware systems and software packages will be known to those skilled in the art.

[00169] Another approach provided by this invention, is the computational screening of small molecule databases for chemical entities or compounds that can bind in whole, or in part, to a DPPIV binding pocket. In this screening, the quality of fit of such entities to the binding site may be judged either by shape complementarities or by estimated interaction energy [E. C. Meng et al., J. Comp. Chem., 13, 505-524 (1992)].

[00170] According to another embodiment, the invention provides compounds that associate with a DPPIV -like binding pocket produced or identified by various methods set forth above.

[00171] The structure coordinates set forth in Figure 3 can also be used to aid in obtaining structural information about another crystallized molecule or molecular complex.

This may be achieved by any of a number of well-known techniques, including molecular replacement.

[00172] For example, a method is also provided for utilizing molecular replacement to obtain structural information about a protein whose structure is unknown comprising the steps of: generating an X-ray diffraction pattern of a crystal of the protein whose structure is unknown; generating a three-dimensional electron density map of the protein whose structure is unknown from the X-ray diffraction pattern by using at least a portion of the structure coordinates set forth in Figure 3 as a molecular replacement model.

[00173] By using molecular replacement, all or part of the structure coordinates of the DPPIV provided by this invention (and set forth in Figure 3) can be used to determine the structure of another crystallized molecule or molecular complex more quickly and efficiently than attempting an *ab initio* structure determination. One particular use includes use with other S9 proteases. Molecular replacement provides an accurate estimation of the phases for an unknown structure. Phases are a factor in equations used to solve crystal structures that cannot be determined directly. Obtaining accurate values for the phases, by methods other than molecular replacement, is a time-consuming process that involves iterative cycles of approximations and refinements and greatly hinders the solution of crystal structures. However, when the crystal structure of a protein containing at least a homologous portion has been solved, the phases from the known structure provide a satisfactory estimate of the phases for the unknown structure.

[00174] Thus, this method involves generating a preliminary model of a molecule or molecular complex whose structure coordinates are unknown, by orienting and positioning the relevant portion of DPPIV according to Figure 3 within the unit cell of the crystal of the unknown molecule or molecular complex so as best to account for the observed X-ray diffraction pattern of the crystal of the molecule or molecular complex whose structure is unknown. Phases can then be calculated from this model and combined with the observed X-ray diffraction pattern amplitudes to generate an electron density map of the structure whose coordinates are unknown. This, in turn, can be subjected to any well-known model building and structure refinement techniques to provide a final, accurate structure of the unknown crystallized molecule or molecular complex [E. Lattman, "Use of the Rotation and Translation

Functions", in Meth. Enzymol., 115, pp. 55-77 (1985); M. G. Rossmann, ed., "The Molecular Replacement Method", Int. Sci. Rev. Ser., No. 13, Gordon & Breach, New York (1972)].

[00175] The structure of any portion of any crystallized molecule or molecular complex that is sufficiently homologous to any portion of DPPIV can be resolved by this method.

[00176] In one embodiment, the method of molecular replacement is utilized to obtain structural information about the present invention and any other DPPIV-like molecule. The structure coordinates of DPPIV, as provided by this invention, are particularly useful in solving the structure of other isoforms of DPPIV or DPPIV complexes.

[00177] The structure coordinates of DPPIV as provided by this invention are useful in solving the structure of DPPIV variants that have amino acid substitutions, additions and/or deletions (referred to collectively as "DPPIV mutants", as compared to naturally occurring DPPIV). These DPPIV mutants may optionally be crystallized in co-complex with a ligand, such as an inhibitor, substrate analogue or a suicide substrate. The crystal structures of a series of such complexes may then be solved by molecular replacement and compared with that of DPPIV. Potential sites for modification within the various binding sites of the enzyme may thus be identified. This information provides an additional tool for determining the most efficient binding interactions such as, for example, increased hydrophobic interactions, between DPPIV and a ligand. It is noted that the ligand may be the protein's natural ligand or may be a potential agonist or antagonist of a protein.

[00178] All of the complexes referred to above may be studied using well-known X-ray diffraction techniques and may be refined versus 1.5-3Å resolution X-ray data to an R value of about 0.22 or less using computer software, such as X-PLOR [Yale University, COPYRGT.1992, distributed by Molecular Simulations, Inc.; see, e.g., Blundell & Johnson, *supra*; Meth. Enzymol., Vol. 114 & 115, H. W. Wyckoff et al., eds., Academic Press (1985)]. This information may thus be used to optimize known DPPIV inhibitors, and more importantly, to design new DPPIV inhibitors.

[00179] The structure coordinates described above may also be used to derive the dihedral angles, phi and psi, that define the conformation of the amino acids in the protein backbone. As will be understood by those skilled in the art, the phi_n angle refers to the rotation around the bond between the alpha-carbon and the nitrogen, and the psi_n angle refers to the

rotation around the bond between the carbonyl carbon and the alpha-carbon. The subscript "n" identifies the amino acid whose conformation is being described [for a general reference, see Blundell and Johnson, Protein Crystallography, Academic Press, London, 1976].

9. Uses of the Crystal and Diffraction Pattern of DPPIV

[00180] Crystals, crystallization conditions and the diffraction pattern of DPPIV that can be generated from the crystals also have a range of uses. One particular use relates to screening entities that are not known ligands of DPPIV for their ability to bind to DPPIV. For example, with the availability of crystallization conditions, crystals and diffraction patterns of DPPIV provided according to the present invention, it is possible to take a crystal of DPPIV; expose the crystal to one or more entities that may be a ligand of DPPIV; and determine whether a ligand/DPPIV complex is formed. The crystals of DPPIV may be exposed to potential ligands by various methods, including but not limited to, soaking a crystal in a solution of one or more potential ligands or co-crystallizing DPPIV in the presence of one or more potential ligands. Given the structure coordinates provided herein, once a ligand complex is formed, the structure coordinates can be used as a model in molecular replacement in order to determine the structure of the ligand complex.

[00181] Once one or more ligands are identified, structural information from the ligand/ DPPIV complex(es) may be used to design new ligands that bind tighter, bind more specifically, have better biological activity or have better safety profile than known ligands.

[00182] In one embodiment, a method is provided for identifying a ligand that binds to DPPIV comprising: (a) attempting to crystallize a protein that comprises a sequence with 70, 80, 90, 95% or greater identity with SEQ. ID No. 1 in the presence of one or more entities; (b) if crystals of the protein are obtained in step (a), obtaining an X-ray diffraction pattern of the protein crystal; and (c) determining whether a ligand/protein complex was formed by comparing an X-ray diffraction pattern of a crystal of the protein formed in the absence of the one or more entities to the crystal formed in the presence of the one or more entities.

[00183] In another embodiment, a method is provided for identifying a ligand that binds to DPPIV comprising: soaking a crystal of a protein that comprises a sequence with 70, 80, 90, 95% or greater identity with SEQ. ID No. 1 with one or more entities; determining whether a

ligand/protein complex was formed by comparing an X-ray diffraction pattern of a crystal of the protein that has not been soaked with the one or more entities to the crystal that has been soaked with the one or more entities.

[00184] Optionally, the method may further comprise converting the diffraction patterns into electron density maps using phases of the protein crystal and comparing the electron density maps.

[00185] Libraries of "shape-diverse" compounds may optionally be used to allow direct identification of the ligand-receptor complex even when the ligand is exposed as part of a mixture. According to this variation, the need for time-consuming de-convolution of a hit from the mixture is avoided. More specifically, the calculated electron density function reveals the binding event, identifies the bound compound and provides a detailed 3-D structure of the ligand-receptor complex. Once a hit is found, one may optionally also screen a number of analogs or derivatives of the hit for tighter binding or better biological activity by traditional screening methods. The hit and information about the structure of the target may also be used to develop analogs or derivatives with tighter binding or better biological activity. It is noted that the ligand-DPPIV complex may optionally be exposed to additional iterations of potential ligands so that two or more hits can be linked together to make a more potent ligand. Screening for potential ligands by co-crystallization and/or soaking is further described in U.S. Patent No. 6,297,021, which is incorporated herein by reference.

EXAMPLES

Example 1. Expression and Purification of DPPIV

[00186] This example describes the expression of DPPIV. It should be noted that a variety of other expression systems and hosts are also suitable for the expression of DPPIV, as would be readily appreciated by one of skill in the art.

[00187] The portion of the gene encoding residues 51-778 (from SEQ. ID No. 1), which corresponds to the extracellular portion of human DPPIV, was isolated by PCR from spleen cDNA and cloned into the *Bam*H I and *Hin*d III sites of a modified pFastBacHTb vector. This vector encodes a baculovirus glycoprotein gp67 signal peptide sequence followed by a 6x-histidine tag sequence followed by the DPPIV sequence. Expression in this vector allowed for

the production of secreted recombinant DPPIV with a 6x-histidine tag, the sequence of which is shown in Figure 1 (6x-histidine tag sequence underlined) (SEQ. ID No. 3).

[00188] Recombinant baculovirus genomic DNAs incorporating the DPPIV cDNA sequences were generated by transposition using the Bac-to-Bac system (Gibco-BRL). Infectious extracellular virus particles were obtained by transfection of a 2 ml adherent culture of *Spodoptera frugiperda Sf9* insect cells with the recombinant viral genomic DNA. Growth in ESF 921 protein free medium (Expression Systems) was for 3 days at 27°C. The resulting passage 1 viral supernatant was used to obtain passage 2 high titer viral stock (HTS) by infection of a 2 ml adherent culture of *Spodoptera frugiperda Sf9* insect cells grown under similar conditions. Passage 2 HTS was used in turn to infect a 100 ml suspension culture of *Spodoptera frugiperda Sf9* insect cells in order to generate passage 3 HTS. The production of recombinant DPPIV proteins was carried out by using the passage 3 HTS at a multiplicity of infection (MOI) of approximately 5 to infect 0.5-5 liter cultures of *Trichoplusia ni* Hi5 insect cells (InVitrogen) at a cell density of (1.5-8)x10^6 cells/ml (grown in ESF 921 protein free medium). Infected cell cultures were grown in both shake flasks and in Wave Bioreactors (Wave Biotech) for 48 hours at 27°C prior to harvest. In some instances infected cultures of *Spodoptera frugiperda Sf9* insect cells were used to produce recombinant DPPIV under similar conditions. Following harvest, the cell cultures were centrifuged to pellet whole cells.

[00189] The secreted glycosylated recombinant protein was isolated from the cell culture medium by diafiltration using cross-flow ultrafiltration, followed by passage over a nickel chelate resin and optionally polished by size exclusion chromatography.

[00190] In a typical batch prep, 5 L of cell culture supernatant was concentrated to 0.1 L on a 10 kDa NMWCO Omega Ultrasette (Pall Life Sciences) using a Masterflex L/S pump fitted with PharMed #15 tubing at a cross flow of approximately 1 L/minute and an inlet feed pressure of 1.5 to 2.0 bar, generating an initial permeate flow of up to 70 ml/minute. The retentate was diluted two to three fold by adding 25 mM Tris/HCl pH 7.9, 0.4 M NaCl and reconcentrated to 0.1 L. This process was repeated at least twice, after which the concentrate was quantitavely removed from the system, centrifuged when necessary (15 minutes at 4000 rpm in an Allegra (Beckman) centrifuge) and added to approximately 8 ml of a preconditioned 50 % slurry of Probond (Invitrogen) divided over three or four 50ml conical tubes. The tubes

were rotated for at least 1 hour, after which the resin was washed with 10 resin volumes of 50 mM Potassium Phosphate pH 7.9, 0.4 M NaCl, 0.25 mM TCEP. The resin is poured into 1 cm ID glass columns (Omnifit) and washed with 50 column volumes of 50 mM Potassium Phosphate pH 7.9, 0.4 M NaCl, 20 mM imidazole, 0.25 mM TCEP. After a wash with 5 column volumes of 50 mM Tris pH 7.9, 0.4 M NaCl, 0.25 mM TCEP, the product is eluted with 4 column volumes of 50 mM Tris pH 7.9, 0.4 M NaCl, 200 mM imidazole, 0.25 mM TCEP.

[00191] It is noted that the polyhistidine tags may optionally be removed; however in this instance, the polyhistidine tag was left as a fusion. It is also noted that for the purification of non-secreted proteins, leupeptin is added to all the buffers used during the IMAC process at 1 mg/L and that for simplicity reasons the same is sometimes done when purifying DPPIV.

[00192] After concentrating to 7.5 mg/ml or higher by centrifugal ultrafiltration (10 kDa NMWCO, VivaScience), DPPIV was purified over a BioSep Sec S3000 column (200mm x 21.2 mm, Phenomenex) at 8 ml/minute to remove oligomeric forms. The column was set up in a Summit HPLC system (Dionex) managed by Chromeleon software (Dionex) and equilibrated with 25 mM Tris pH 7.6, 150 to 250mM NaCl (optionally with 0.25 mM TCEP and 1 mM EDTA). In cases when the size exclusion step was omitted, centrifugal ultrafiltration (10 kDa NMWCO) was used for the buffer exchange to the required formulation buffer. The process was carried out at 2–10°C and DPPIV was stored at the same temperature. For long-term storage, it was kept at –80°C. The purity of DPPIV was estimated by SDS-PAGE and IEF to be at least 95%. Glycosylation was confirmed by a molecular mass shift, determined by SDS-PAGE, following Endo-F1 enzyme treatment, and by carbohydrate analysis.

[00193] DPPIV with seleno-L-methionine substitution was prepared as follows: two 5 liter Wave Bioreactor cultures of *Trichoplusia ni* Hi5 insect cells in ESF 921 Protein Free medium were infected and grown for 16 hours at 27°C. At that time the cells were pelleted by centrifugation at 480*g* and 20°C for 15 minutes. The supernatant was discarded and the cells resuspended in 2x 5 liters of ESF 921 Protein Free Methionine-Free medium (Expression Systems). The resuspended cells were placed in two new 5 liter Wave Bioreactors and growth continued for 4h at 27°C. Seleno-L-methionine (prepared as a 25 mg/ml solution in water and sterile-filtered) was then added to each culture to a final concentration of 50 mg/l. Cell growth

was continued for a further 48h prior to harvest. Purification of the protein was as described above and included the size exclusion chromatography step. Mass spectrogram peptide analysis was used to estimate the seleno-L-methionine substitution of methionine residues at approximately 34%.

Example 2. Crystallization of DPPIV

[00194] This example describes the crystallization of DPPIV. It is noted that the precise crystallization conditions used may be further varied, for example by performing a fine screen based on these crystallization conditions.

[00195] Crystals were obtained after an extensive and broad screen of conditions, followed by optimization. Diffraction quality crystals were grown in 100nL sitting droplets using the vapor diffusion method. 50nL comprising the apo DPPIV complex (between 8 and 30 mg/ml) was mixed with 50nL from a reservoir solution (100μL) comprising: 0.1M Tris-HCl, pH=7.5; 27% MPEG 2000; and 0.35M sarcosine/10% xylitrol. The resulting solution was incubated over a period of two weeks at 4°C.

[00196] Crystals typically appeared after 3-5 days and grew to a maximum size within 7-10 days. Single crystals were transferred, briefly, into a cryoprotecting solution containing the reservoir solution supplemented with 25% v/v ethylene glycol. Crystals were then flash frozen by immersion in liquid nitrogen and then stored under liquid nitrogen. A crystal of apo DPPIV produced as described is illustrated in Figure 2.

[00197] While the present invention is disclosed with reference to certain embodiments and examples detailed above, it is to be understood that these embodiments and examples are intended to be illustrative rather than limiting, as it is contemplated that modifications will readily occur to those skilled in the art, which modifications are intended to be within the scope of the invention and the appended claims. All patents, papers, and books cited in this application are incorporated herein in their entirety.